Bulletin KRASEC. Phys. & Math. Sci, 2014, V. 9, №. 2, pp. 70-78. ISSN 2313-0156

DOI: 10.18454/2313-0156-2014-9-2-70-78

MSC 65C20

COMPARISON OF THE SPARSE APPROXIMATION METHODS BASED ON ITS USE TO GEOACOUSTIC EMISSION SIGNALS

O.O. Lukovenkova[1, 2]

[1] Institute of Cosmophysical Researches and Radio Wave Propagation Far-Eastern Branch, Russian Academy of Sciences, 684034, Kamchatskiy Kray, Paratunka, Mirnaya st., 7, Russia

[2] Vitus Bering Kamchatka State University, 683031, Petropavlovsk-Kamchatsky, Pogranichnaya st., 4, Russia

E-mail: o.o.lukovenkova@yandex.ru

The paper is devoted to the comparative analysis of some sparse approximation methods. The first part of the paper describes general sparse approximation problem and two main approaches solved it. Classification of testing pursuit algorithm is illustrated. Features of the methods application to geoacoustic emission signals are considered in the second part. The sparseness, accuracy and runtime of described pursuit algorithms are compared.

Key words: matching pursuit, basis pursuit, geoacoustic emission

Introduction

Since 1999, investigation of geoacoustic emission (GAE) signals has been carried out at different stages of seismic activity at the Institute of Cosmophysical Research and Radio Wave Propagation (IKIR) FEB RAS. A typical GAE signal is a series of relaxation pulses with shock excitation, amplitude of 0.1 – 1 Pa, filling frequency from the units to the first tens of kHz (Fig. 1) [1].

Signal registration is carried out continuously with the sampling frequency of 48 kHz which does not allow manual processing. As a rule, classical methods of frequency-time analysis are applied for the analysis of pulse nature signals. However, during the last years, methods of sparse approximation become more popular. The paper compares sparse approximation algorithms in accuracy, sparseness of the obtained solution and time of execution on the example of GAE signals.

Lukovenkova Olga Olegovna – Assistent of Dept. Informatics, Vitus Bering Kamchatka State University, Postgraduate Student, Institute of Cosmophysical Research and Radio Wave Propagation FEB RAS.

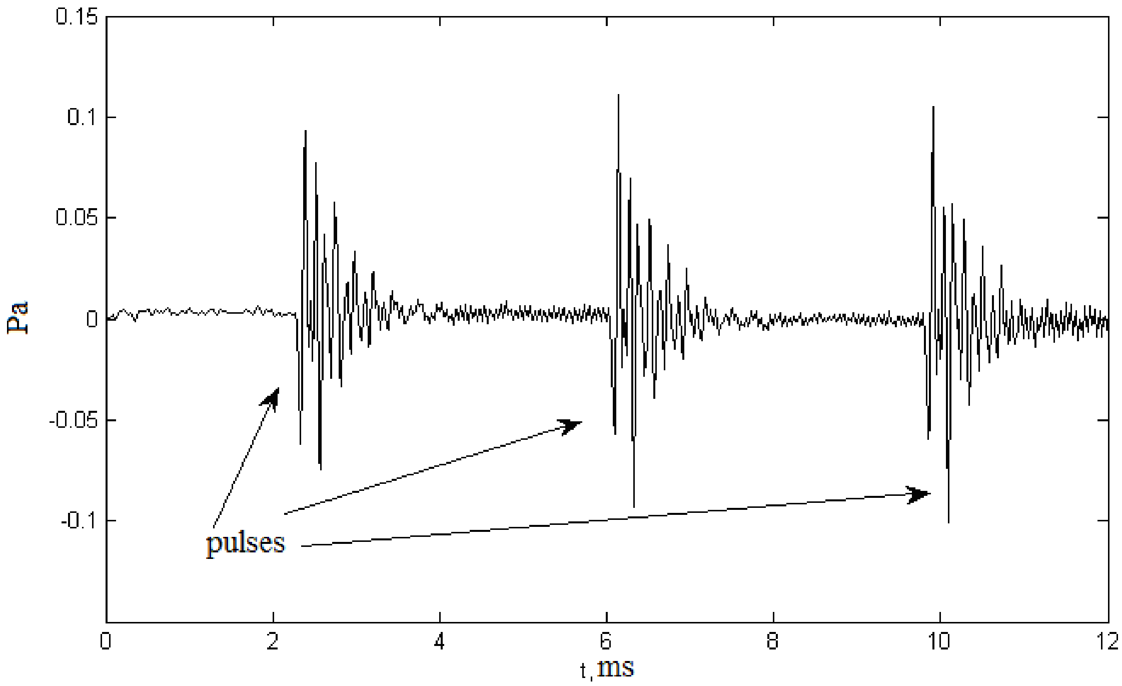

Fig. 1. GAE signal

Sparse approximation

Signal approximation is a problem of signal decomposition in some function set (function dictionary):

$$f(t) = \sum_{m=0}^{N-1} a_m g_m(t) + R_N, \|R_N\| \to \min,$$

where $f(t)$ is the signal under the investigation, $g_m(t)$ is the element (atom) of the dictionary $D = \{g_m(t), \quad \|g_m\| = 1\}$, a_m – are the decomposition coefficients, N is the number of decomposition elements, R_N is the approximation error.

Sparse approximation assumes building of a signal model, containing the least number of elements, i.e.

$$f(t) = \sum_{m=0}^{N-1} a_m g_m(t) + R_N, \|R_N\| \to \min, \|a\|_0 \to \min,$$

where $\|\cdot\|_0$ is the pseudonorm (L_0-norm), equal to the number of nonvanishing terms of a vector.

One of the main advantages of sparse approximation is the possibility to build signal decomposition simultaneously containing the least number of elements and minimizing the error in redundant unorthogonal dictionary, in general case. A redundant dictionary contains the number of atoms which significantly exceeds the dimensionality of the initial signal. However, such type of problems of search for an optimum basis of decomposition is very complicated for computation and may not be solved for the polynomial time.

There are two different approaches to the sparse approximation of signals. The both make the computation of the given problem not so complicated and find an effective but not an optimum solution:

1) Basis Pursuit (BP). The essence of this approach, suggested by Chen S.S. and Donoho D.L. [2], comes to the change of the computationally complicated problem of L_0-approximation by an easier problem of L_1-approximation.

$$D \cdot a = f, \|a\|_1 \to \min,$$

where

$$\|a\|_1 = \sum_{m=0}^{N-1} |a_m|.$$

Minimization of L_1-norm eliminates energy f dissipation in dictionary D atoms, thus, reducing the number of elements of the desired decomposition.

The basis pursuit problem may be simplified by the change of L_1-norm minimization by its limitation [2].

$$f(t) = \sum_{m=0}^{N-1} a_m g_m(t) + R_N, \|R_N\| \to \min, \|a\|_1 < \lambda.$$

We should note that basis pursuit is an optimization principle rather than a concrete algorithm for solution of a problem. The basis pursuit problem may be solved by the reduction to the problem of linear programming [2] or by one of optimization methods [3],[4] refining the given initial approximation at each iteration.

1) Matching Pursuit (MP) was suggested by Mallat S. and Zhang Z. [5]. The essence of the algorithm comes to the iterative process of search for dictionary elements minimizing the approximation error at every step.

$$\begin{cases} R^0 f = f \\ s = \arg\left[\min_m \left|\left\langle g_m, R^N f \right\rangle\right|\right] \\ R^{N+1} f = R^N f - \left\langle g_s, R^N f \right\rangle g_s \end{cases}.$$

On the basis of the matching pursuit method, an orthogonal matching pursuit (OMP) method was developed. The main difference of it from the classical realization is the orthogonal basis, minimizing the approximation error [6].

$$\begin{cases} R^0 f = f, \qquad U = \emptyset \\ s = \arg\left[\min_{m \notin U} \left|\left\langle g_m, R^N f \right\rangle\right|\right] \\ U = U \cup s \\ a^N = D_U^+ f \\ R^{N+1} f = f - D \cdot a^N \end{cases}.$$

Classification of approximation algorithms for signals, tested on GAE signals, is illustrated in Fig. 2.

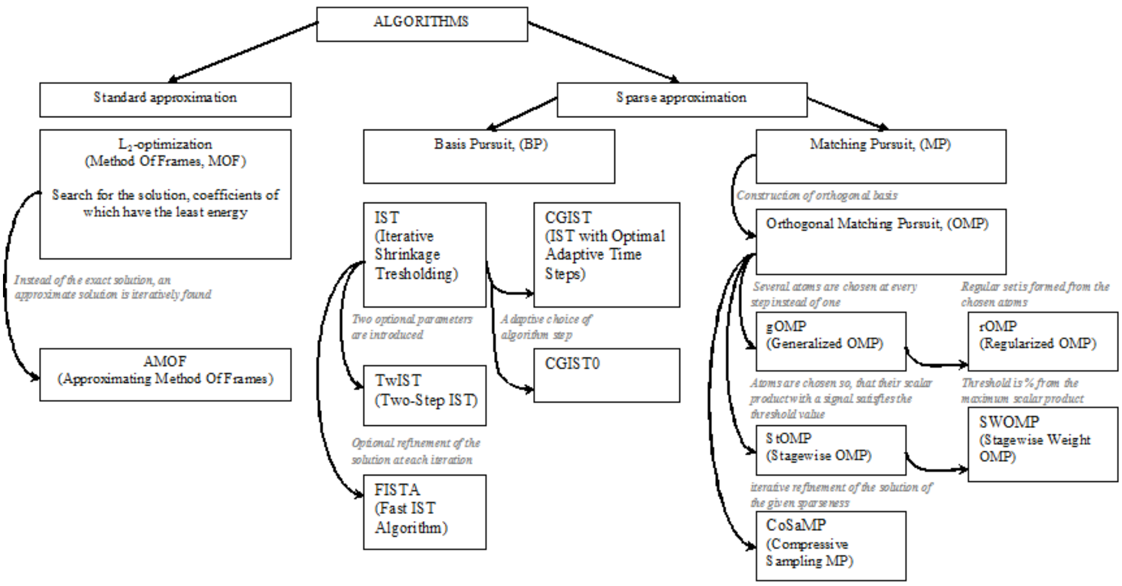

Fig. 2. Classification of signal approximation algorithms

Comparison of sparse approximation methods on GAE signals

To compare the algorithms of sparse approximation, a sample was formed which contained 100 clear pulses, distinguished from GAE signals, with the amplitude of 0.02 – 0.05 Pa, filling frequency of 5-10 kHz (Fig. 3), and duration of 8 ms. Preliminary processing of signals included filtration in the frequency range of 1-24 kHz and amplitude normalizing.

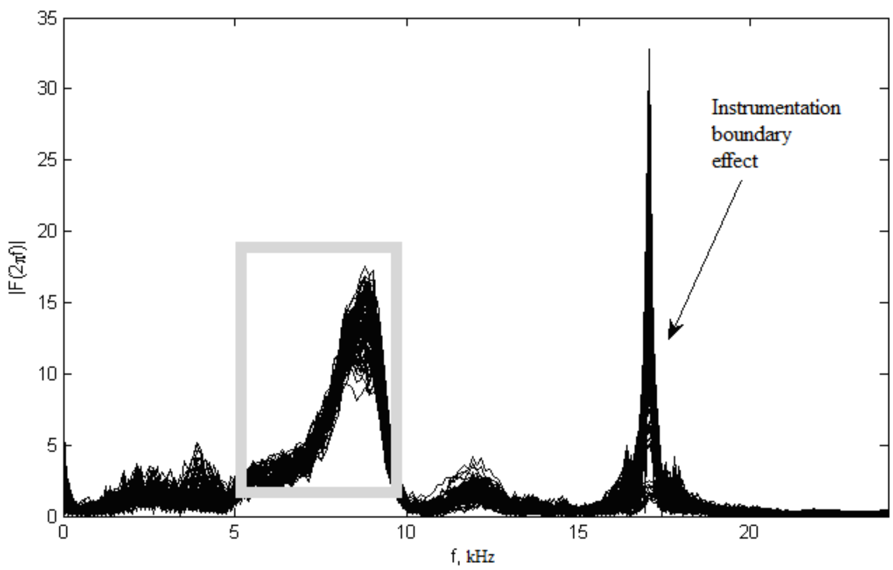

Fig. 3. Spectrum of the chosen pulses

Choice of the dictionary D is an important problem which determines the quality of further analysis. The previous papers showed that the most effective dictionary for geoacoustic signal approximation is the dictionary formed from Berlage modulated functions, since Berlage pulses have similar structure with geoacoustic emission elementary pulses [7]-[9]. In the course of the experiments, a dictionary was chosen which gave an appropriate accuracy of approximation. The dictionary, applied in the experiment, contained 2460 Berlage functions with the following parameters: duration of 3.8 ms, envelope maximum at 152 – 0.95 ms from the pulse beginning, filling sinusoid frequency from 4,5 to 10,5 kHz (Fig. 4).

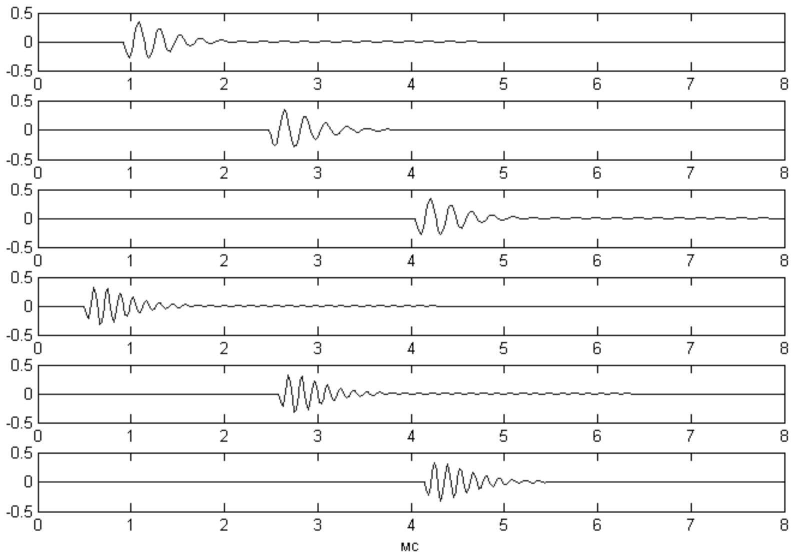

Fig. 4. Examples of dictionary atoms

For each signal under the investigation, a classical approximation (all dictionary atoms were included into signal decomposition) was constructed obtained by solving the L_2-optimization problem. In the result of the analysis of the obtained decomposition coefficient vectors, it was determined that in every of the 100 solutions, all the 2460 coefficients were nonvanishing. However, in average, only 991 coefficients in the decomposition had values exceeding 1% from the maximum, and only 294 coefficients exceeded the threshold of 5% (Fig. 5).

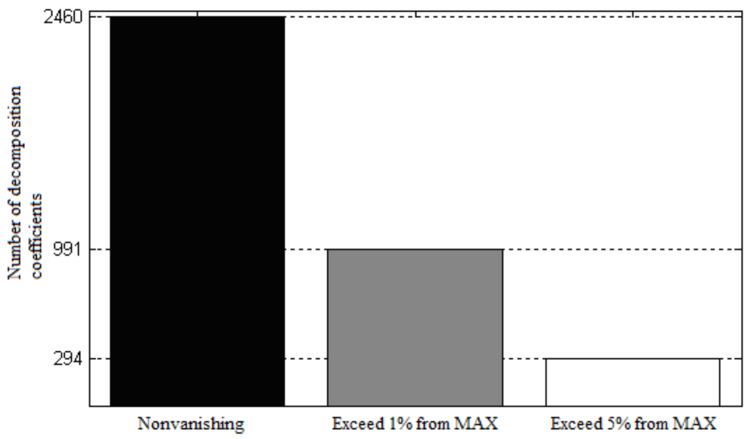

Fig. 5. Results of L_2-optimization

Thus, the correlation of more than a half of the dictionary atoms with the signal was insignificant, so the obtained decomposition was redundant. Redundancy of the presentation may be avoided if sparse approximation of the given data is executed on the defined dictionary.

Applying the sparse approximation method, one should remember, that in comparison to matching pursuit, it is impossible in the basis pursuit algorithms to define the exact limitation on L_0-norm of the desired coefficient vector and sparseness degree is regulated by a control parameter μ, the higher its value is, the less is the L_0-norm of decomposition coefficients and larger the error [4].

$$\mu\,\|a\|_1 + \frac{1}{2}\left\|R^N f\right\|^2 \to \min.$$

The choice of the proper values for the control parameter and algorithm step, which ensure the required relation «sparseness-error», is a complicated problem solved only by experimental selection (Fig. 6).

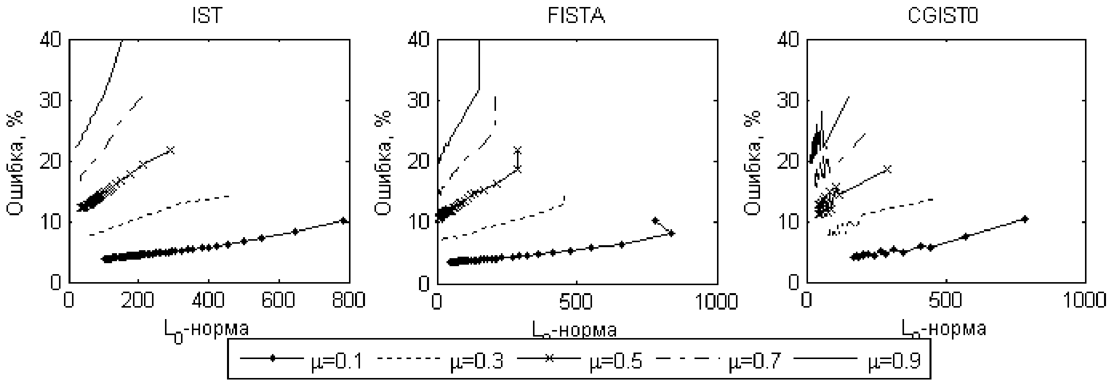

Fig. 6. Influence of parameter μ on the solution of basis pursuit problem

In most of the matching pursuit algorithms, except for StOMP and SWOMP, L_0-norm of coefficient vectors either depends on the number of iterations (it coincide in the case with MP and OMP) or is directly indicated as an input parameter of the algorithm (CoSaMP) that is more convenient for the researcher, if the priority is given to the solution sparseness.

Fig. 7 shows the graphs of the dependence for average approximation error of the investigated signals on the mean value of coefficient vector L_0-norm for different algorithms of sparse approximation. The values 0.5 and 0.7 were experimentally selected as the control parameters μ for basis pursuit algorithms. They gave the proper error level and solution sparseness. TwIST pursuit algorithm is applied for the dictionaries satisfying the condition $0 < k \leq \lambda_{\min}(D^T D) \leq 1$, for the defined dictionary $\lambda_{\min}(D^T D) < 0$, so this algorithm was excluded from the list of tested ones.

From the graphs in Fig. 7(a) and 7(b), the differences in the dynamics of error decrease for matching and basis pursuit methods are clear. In the first case, the error gradually decreases with the increase of the number of decomposition elements, in the second case it decreases with the decrease of element number.

If every algorithm is presented as an ellipse with semiaxes, corresponding to confidence intervals of resultant L_0-norm and error average values, then the efficiency of application

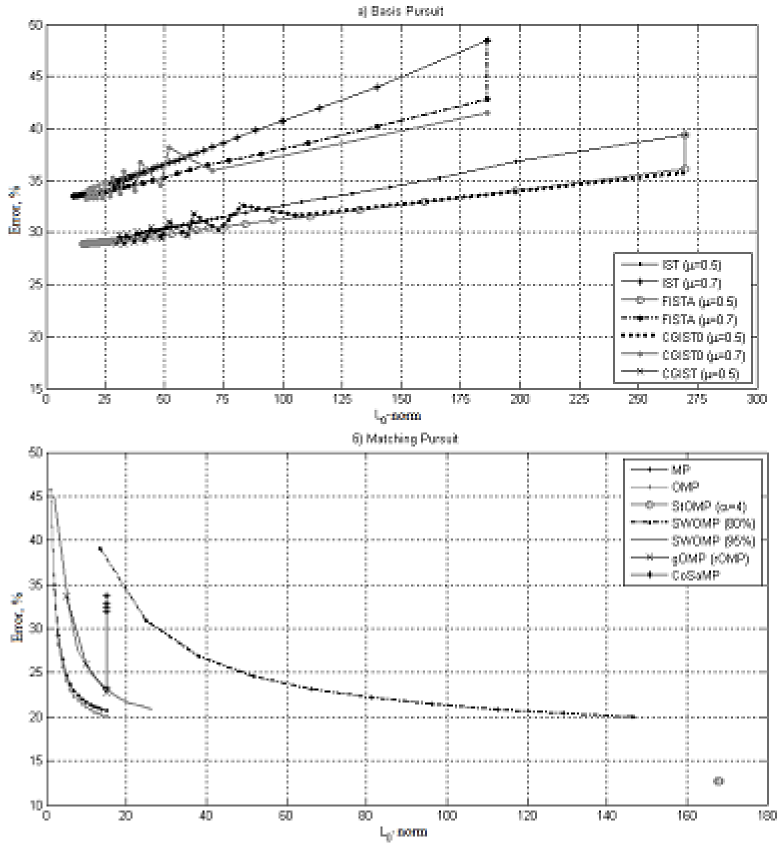

Fig. 7. Dependence of error decrease on solution sparseness

of the algorithm for the investigated signals and the defined dictionary may be evaluated by the arrangement of this ellipse on the coordinate plate (Fig. 8).

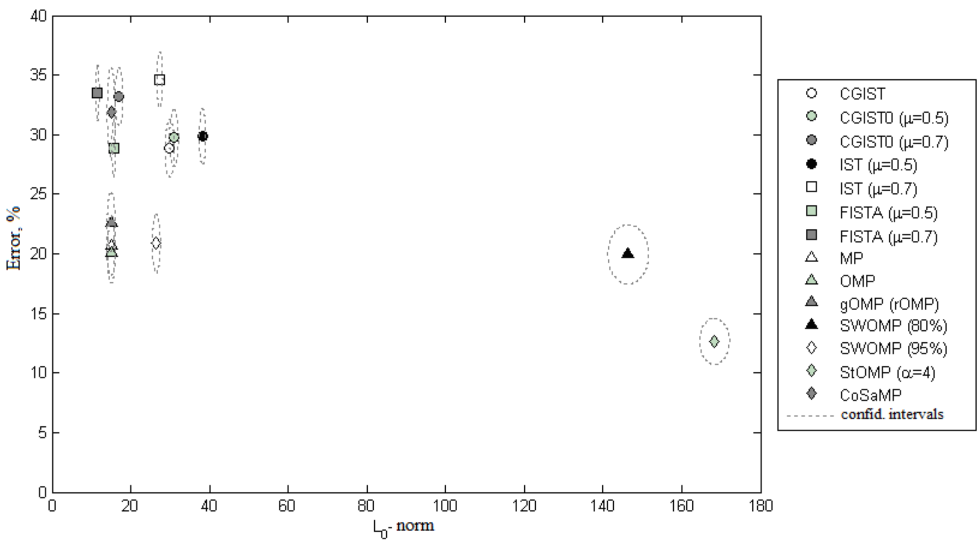

Fig. 8. The relation «sparseness-error» for sparse approximation algorithms

Table shows the time for execution of 20 algorithm iterations, sparseness and error averaged for 100 pulses. The computations were carried out on a PC with the following characteristics: Intel Core i5-3210M processor (2.50 GHz), main memory size 4 Gb.

Table

Pursuit algorithm characteristics

Method	**Execution time for 20 iter., sec**	**Res. L_0**	**Res. error**
OMP	0.8644	15	20%
MP	0.7927	15	21%
gOMP (rOMP)	time depends on the parameters 10 it. with 2 atoms: 0.4693 (0.5191) 4 it. with 5 atoms 0.2423 (0.2832) 2 it. with 10 atoms: 0.1651 (0.2119)	15	23%
FISTA (μ=0.5)	1.3582	16	29%
SWOMP (95%)	0.9770	26	21%
CoSaMP	0.1310	15	32%
FISTA (μ=0.7)	1.3582	12	33%
CGIST0 (μ=0.7)	15.6585	17	33%
CGIST (μ=0.5)	27.8995	30	29%
CGIST0 (μ=0.5)	15.6585	31	30%
IST (μ=0.7)	1.3509	27	35%
IST (μ=0.5)	1.3509	38	30%
SWOMP (80%)	2.0469	146	20%
StOMP (α=4)	1.3351	168	13%

According to the graph, shown in Fig. 8, and Table data, MP, OMP, gOMP matching pursuit methods are the most effective for the relation «sparseness-error». In the course of testing of rOMP, it was determined that for the investigated signals, a set of atoms, chosen for each iteration, is already regular, so approximation solutions, obtained by rOMP and gOMP, are identical. From the basis pursuit methods, FISTA gave the best results. Comparison of execution times for the algorithms showed that matching pursuit algorithms are faster than the basis pursuit ones, as long as, in practice, 15-20 iterations of matching pursuit algorithms and 50-70 iterations of basic pursuit algorithms are needed to achieve an optimum result.

Conclusion

In the result of the study, we may conclude that the solutions obtained by matching pursuit algorithms, MP, OMP, and gOMP, have better sparseness and higher accuracy in comparison to the solutions obtained by other considered algorithms. The algorithms are less complicated and, consequently, their application in GAE signal analysis is reasonable and effective. The choice between the three algorithms is determined by the requirements of a researcher: the fastest one is gOMP, the most accurate one is OMP, MP is the «golden mean».

References

1. Marapulec Yu.V. Shevcov B.M. *Mezomasshtabnaya akusticheskaya 'emissiya* [Mesoscale acoustic emission]. Vladivostok, Dal'nauka Publ., 2012.

2. Chen S. S., Donoho D. L., Saunders M. A. Atomic decomposition by basis pursuit. *SIAM Journal on Scientic Computing*, 1998, vol. 20, no. 1, pp. 33–61.

3. Beck A. and Teboulle M. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *Siam J. Imaging Sciences*, 2009, vol. 2, no. 1, pp. 183–202.

4. Goldstein T., Setzer S. High-order methods for basis pursuit. Los Angeles, Preprint University of California, 2010.

5. Mallat S., Zhang Z. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 1993, 41(12), pp. 3397-3415.

6. Tropp J. A. and Gilbert A. C. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions Information Theory*, 2007, vol. 53, no. 12, pp. 4655–4666.

7. Marapulec Yu.V., Tristanov A.B. Primenenie metoda razrezhennoj approksimacii v zadachah analiza signalov geoakusticheskoj 'emissii [Application of the sparse approximation for the analysis of signals geoacoustical emission]. *Cifrovaya obrabotka signalov – Digital signal processing*, 2011, no. 2, pp. 13-17.

8. Marapulec Yu.V., Tristanov A.B. Razrezhennaya approksimaciya akusticheskih vremennyh ryadov s ispol'zovaniem chastotno-vremennogo slovarya Berlage [Sparse approximation of the acoustic time series with time-frequency vocabulary Berlage]. Trudy Rossijskogo nauchno-tehnicheskogo obschestva radiotehniki, 'elektroniki i svyazi im. A.S. Popova. Seriya: Cifrovaya obrabotka signalov i ee primenenie [Proc. of the Russian Scientific and Technical Society of Radio Engineering, Electronics and Communication named after AS Popova]. Moscow, 2012, vol. 1, pp. 91-94.

9. Afanas'eva A.A., Lukovenkova O.O. Metody obnaruzheniya impul'sov geoakusticheskoj 'emissii na osnove algoritmov razrezhennoj approksimacii i klasterizacii [Methods of detection pulses geoacoustical emission based on sparse approximation algorithms and clustering]. *Vestnik KRAUNC. Fiziko-matematicheskie nauki – Bulletin of the Kamchatka Regional Association «Education-Scientific Center». Physical & Mathematical Sciences*, 2013, vol. 7, no. 2, pp. 68-73.

Original article submitted: 29.11.2014